# EXHIBIT W

# Nonparametric Estimation of a Survivorship Function with Doubly Censored Data

BRUCE W. TURNBULL *

A simple iterative procedure is proposed for obtaining estimates of a response time distribution when some of the data are censored on the left and some on the right. The procedure is based on the product-limit method of Kaplan and Meier [15], and it also uses the idea of self-consistency due to Efron [8]. Under fairly general assumptions, the method is shown to yield unique consistent maximum likelihood estimators. Asymptotic expressions for their variances and covariances are derived and an extension to the case of arbitrary censoring is suggested.

## 1. INTRODUCTION AND SUMMARY

A common problem in statistical analysis is the determination of the distribution of the time, $T$, taken for an event of interest to occur. For instance, in medical follow-up studies, the event of interest is the relapse or death of a patient; and in life-testing it is the time to failure of an item that is under investigation. In this article all such events will be termed "deaths," and thus the problem is to estimate the lifetime distribution, that is $F(t) = \text{Prob}\,(T \leq t)$ for $t \geq 0$.

In a sample of $N$ observations $T_1, T_2, \cdots, T_N$, where each lifetime $T_i$ is observed precisely, the natural estimate is the sample distribution function $\hat{F}(t)$, which is the proportion of observations with values less than or equal to the argument $t$. In this article we shall consider the situation where not all the $T_i$ are observed exactly but some are censored on the right and some on the left. For each item $i$ $(1 \leq i \leq N)$, we assume that there are limits of observation $L_i$ and $U_i$ (with $L_i \leq U_i$), which are either fixed constants or random variables independent of the $\{T_i\}$. Thus $(L_i, U_i)$ is a "window" of observation and the recorded information is:

$$X_i = \max\,[\min\,(T_i, U_i), L_i].$$

Also, for each item $i$, it is known whether $X_i = L_i$ (i.e., $T_i \leqslant L_i$ and the item is left censored or a "late entry"), or $X_i = U_i$ (i.e., $T_i > U_i$ and the item is right censored or a "loss"), or $X_i = T_i$ (i.e., $L_i < T_i \leqslant U_i$). We can denote a loss at time $t$ by the symbol "$>t$," a late entry by "$\leqslant t$," and a precise observation by simply "$t$."

Gehan [10], Mantel [19], and Peto [22] have given examples where double censoring might arise in medical applications. Another example occurred in a recent study of African infant precocity by Leiderman *et al.* [17]. Their purpose was to establish norms for infant development for a community in Kenya in order to make comparisons with known standards in the United States and the United Kingdom. The sample consisted of 65 children born between July 1 and December 31, 1969. Starting in January 1970, each child was tested monthly to see if he had learned to accomplish certain standard tasks (see [4]). Here $T$ would represent the time from birth to first learn to perform a particular task. Late entries occurred when it was found that, at the very first test, some children could already perform the task; whereas losses occurred when some infants were still unsuccessful by the end of the study.

The more common case is when there are no late entries (all $L_i = 0$) and this has been treated extensively in the literature. Often some parametric form for $F$ is assumed such as an exponential, lognormal or Weibull. The method of maximum likelihood in such a situation was first used by Boag [6] and most recently by Herman and Patell [14] and Moeschberger and David [21]. Nonparametric estimates can be obtained by the actuarial method (see, e.g., [5]), or by the product-limit (PL) or reduced-sample (RS) methods described in [15]. Nonparametric two sample tests for comparing two such lifetime distributions have been proposed by Halperin [12], Gilbert [11], Gehan [9], Mantel [18], and Efron [8]. Two sample tests with doubly censored data have been treated by Gehan [10] and Mantel [19].

In this article we treat the estimation problem when there is both left and right censoring. (This should not be confused with the case of right censoring and left truncation, which is discussed in Mantel [18, p. 166]). We assume that there is a natural discrete time scale $0 < t_1 < t_2 < \cdots < t_m$, which would occur, for instance, if items were examined only at discrete times (monthly in the case of the Leiderman study just mentioned). Alternatively, we can assume that the data is grouped and lifetimes are recorded only as belonging to one of the $m$ intervals $(0, t_1]$, $(t_1, t_2]$, $\cdots$, $(t_{m-1}, t_m]$. We let $\delta_i$ be the number of items observed to have died in age period $(t_{i-1}, t_i]$, $\mu_i$ be the number of late entries at age $t_i$, and

* Bruce W. Turnbull is lecturer in statistics, Mathematical Institute, University of Oxford, 24–29 St. Giles, Oxford, OX1 3LB, England. This article was prepared in part under the auspices of Public Health Service Grant USPHS-5T1 GM 25-14, while the author was visiting the Department of Statistics, Stanford University. The author is grateful to Dr. Helena C. Kraemer, who first showed him this problem, and to Professor Bradley Efron for some useful discussions.

© Journal of the American Statistical Association
March 1974, Volume 69 Number 345
Theory and Methods Section

$\lambda_i$ be the number of losses at $t_i$ ($1 \leqslant i \leqslant m$). The situation is illustrated by the following tabulation:

| *Type of observation* | *Age* $t_1$ | $t_2$ | $\cdots$ | $t_m$ |
|---|---|---|---|---|
| Deaths | $\delta_1$ | $\delta_2$ | | $\delta_m$ |
| Losses ($>$) | $\lambda_1$ | $\lambda_2$ | | $\lambda_m$ |
| Late entries ($\leq$) | $\mu_1$ | $\mu_2$ | | $\mu_m$ |

We have made the assumption that the late entries $\mu_i$ all occur at the end of age period $(t_{i-1}, t_i]$ and that the losses $\lambda_i$ all occur at the beginning of $(t_i, t_{i+1}]$. Alternative assumptions are discussed in Section 4.

In Section 2 a simple iterative procedure is proposed to obtain estimates of $P_1, P_2, \cdots, P_m$ where $P_i = P(t_i)$ and $P(t) = 1 - F(t)$ is the survivorship function. These estimates will be denoted by $\hat{P}_1, \hat{P}_2, \cdots, \hat{P}_m$. The procedure is based on the idea of "self-consistency" which is due to Efron [8]. In Section 3, the procedure is shown to yield unique maximum likelihood estimates which are consistent, and expressions for their variances and covariances are given.

## 2. THE SELF-CONSISTENT ESTIMATORS

First consider the $\mu_i$ late entries at age $t_i$. The estimated mean number of these that die in period $(t_{j-1}, t_j]$ for $j \leqslant i$ is $\mu_i\alpha_{ij}$, where $\alpha_{ij} = (\hat{P}_{j-1} - \hat{P}_j)/(1 - \hat{P}_i)$ is an estimate of Prob $[t_{j-1} < T \leqslant t_j | T \leqslant t_i]$. With this in mind, we now consider an "adjusted" problem obtained by replacing each $\delta_j$ by $\delta_j' = \delta_j + \sum_{i=j}^m \mu_i\alpha_{ij}$, leaving each $\lambda_j$ unchanged, and replacing each $\mu_j$ by zero. Since this adjusted data set is singly censored only, we can write down explicitly its PL estimates $\{\hat{P}_i'\}$ say, using the method of Kaplan and Meier [15]. We say that the estimates $\{\hat{P}_i\}$ are *self-consistent* if $\hat{P}_i' = \hat{P}_i$ ($1 \leqslant i \leqslant m$).

Therefore, the problem is to find numbers $1 \geq \hat{P}_1 \geq \hat{P}_2 \geq \cdots \geq \hat{P}_m \geq 0$ which satisfy the following implicit equations:

$$\hat{P}_1 = q_1 \qquad (2.1)$$

$$\hat{P}_j = q_j\hat{P}_{j-1} \quad (j = 2, 3, \cdots, m)$$

where

$$q_j = (n_j' - \delta_j')/n_j', \quad n_j' = \sum_{i=j}^m (\lambda_i + \delta_i'),$$

$$\delta_j' = \delta_j + \sum_{i=j}^m \mu_i\alpha_{ij}$$

and

$$\alpha_{ij} = (\hat{P}_{j-1} - \hat{P}_j)/(1 - \hat{P}_i) \quad \text{for} \quad j \leq i.$$

An iterative method for solving these equations immediately suggests itself:

A. Obtain initial estimates $\{P_i^0, 1 \leqslant i \leqslant m\}$. This set could be any decreasing sequence of $m$ numbers between 1 and 0, but it might be sensible to take the PL estimates obtained by assuming all the $\{\mu_i\}$ to be zero.

B. Form

$$\alpha_{ij}^0 = (P_{j-1}^0 - P_j^0)/(1 - P_i^0), \quad \text{all } j \leq i,$$

and set

$$\delta_j' = \delta_j + \sum_{i=j}^m \mu_i\alpha_{ij}, \quad 1 \leq j \leq m.$$

C. Obtain improved estimates by taking all $\mu_i' = 0$, replacing $\delta_i$ by $\delta_i'$ ($1 \leq i \leq m$) and forming the PL estimates on this "adjusted" data set, i.e.,

$$P_1^1 = 1 - \delta_1'/n_1',$$

and

$$P_j^1 = q_jP_{j-1}^1, \quad (j = 2, 3, \cdots, m);$$

where

$$q_j = (n_j' - \delta_j')/n_j'$$

and

$$n_j' = \sum_{i=j}^m (\lambda_i + \delta_i').$$

D. Return to Step B with the $\{P_j^0\}$ replaced by $\{P_j^1\}$, etc.

E. Stop when the required accuracy has been achieved (e.g., the rule may be to stop when $\max_{1\leq i\leq m} |P_i^\ell - P_i^{\ell-1}| < 0.001$, say).

The procedure is simple to program on a computer and converges fairly rapidly. As a small numerical example, consider the data in the following tabulation:

| *Type of observation* | *Age* $t_1$ | $t_2$ | $t_3$ | $t_4$ |
|---|---|---|---|---|
| Deaths | 12 | 6 | 2 | 3 |
| Losses | 3 | 2 | 0 | 3 |
| Late entries | 2 | 4 | 2 | 5 |

The initial values are:

$$\delta = 12.0, \quad 6.0, \quad 2.0, \quad 3.0$$

$$\mathbf{P}^0 = 0.613, \quad 0.383, \quad 0.287, \quad 0.144.$$

The first iteration yields:

$$\delta' = 19.9, \quad 9.5, \quad 2.8, \quad 3.8$$

$$\mathbf{P}^1 = 0.549, \quad 0.303, \quad 0.214, \quad 0.094.$$

After three iterations, the values settle down on:

$$\delta' = 20.3, \quad 9.3, \quad 2.7, \quad 3.6$$

$$\hat{\mathbf{P}} = 0.538, \quad 0.295, \quad 0.210, \quad 0.095.$$

In Section 3 we will show that $\{\hat{P}_i\}$ are in fact maximum likelihood estimates. It should be noted that there are several special cases when no iteration is needed and explicit estimates can be obtained. First, of course, if all $\mu_i = 0$, then we have single censoring and the PL estimates are available immediately. By reversing the time scale, the PL method can again be applied in the case of left censoring only (all $\lambda_i = 0$). Also if the first tabulation is such that there exists an integer $\ell$ such that $\mu_i = 0$ for all $i > \ell$ and $\lambda_i = 0$ for all $i < \ell$, there is a binomial estimate available for $\hat{P}_\ell$. Then, by "working towards each end of the time scale," the appropriate conditional probabilities can be estimated and the $\{\hat{P}_i\}$ evaluated by the PL method. Finally there is the special case of all $\delta_i = 0$. Ayer *et al.* [1] have derived explicit expressions for the maximum likelihood estimates in this case, and these too have the property of self-consistency.

## 3. THE MAXIMUM LIKELIHOOD DERIVATION

We now state and prove the fundamental theorem of this article.

*Theorem*: If $\delta_i > 0$ ($1 \leq i \leq m$), then the solution $\hat{\mathbf{P}} = (\hat{P}_1, \hat{P}_2, \cdots, \hat{P}_m)$ of (2.1), obtained by the iterative procedure of Section 2, is the unique maximum likelihood estimate of $(P_1, P_2, \cdots, P_m)$.

*Proof*: Under the assumptions made in Section 1, the likelihood function is proportional to:

$$\prod_{j=1}^{m} (P_{j-1} - P_j)^{\delta_j} P_j^{\lambda_j} (1 - P_j)^{\mu_j},$$

where $P_0 = 1$. The log-likelihood $L$ is given by:

$$L = \sum_{j=1}^{m} [\delta_j \log (P_{j-1} - P_j) + \lambda_j \log P_j + \mu_j \log (1 - P_j)].$$

The maximum likelihood estimates will be those values of $\{P_j\}$ which maximize $L$ subject to the condition $1 \geq P_1 \geq \cdots \geq P_m \geq 0$. (This constraint makes the problem similar to those of isotonic regression, see [3].)

First note that if $\lambda_m = 0$, then $L$ is maximized by taking $\hat{P}_m = 0$, and the problem can be treated as one with $m - 1$ periods and $\lambda_{m-1}$ replaced by $\lambda_{m-1} + \delta_m$. In this case $\mu_m$ contributes no information to the estimation of the $\{P_i\}$ and this agrees with intuition. Thus without loss of generality we may assume $\lambda_m > 0$.

Differentiating $L$ with respect to $P_1, P_2, \cdots, P_m$ and setting the derivatives equal to zero, we obtain:

$$\frac{\partial L}{\partial P_j} = -\frac{\delta_j}{P_{j-1} - P_j} + \frac{\delta_{j+1}}{P_j - P_{j+1}} + \frac{\lambda_j}{P_j} - \frac{\mu_j}{1 - P_j} = 0, \quad (j = 1, 2, \cdots, m - 1);$$
$$\frac{\partial L}{\partial P_m} = -\frac{\delta_m}{P_{m-1} - P_m} + \frac{\lambda_m}{P_m} - \frac{\mu_m}{1 - P_m} = 0. \tag{3.1}$$

In Lemma A1 of the appendix, we show that any solution $\{\hat{P}_i\}$ of (2.1) also satisfies the likelihood equations (3.1). Also the estimates obtained in Section 2 clearly satisfy the condition $1 > \hat{P}_1 > \cdots > \hat{P}_m > 0$, since we have assumed $\lambda_m > 0$ and $\delta_i > 0$ $(1 \leq i \leq m)$. Thus the $\{\hat{P}_i\}$ give stationary values of the likelihood function. To show that this is a unique maximum we examine the matrix $\mathbf{D}$ of second derivatives.

Let $D_{ij} = \partial^2 L / \partial P_i \partial P_j$; then

$$D_{ii} = -\frac{\delta_i}{(P_{i-1} - P_i)^2} - \frac{\delta_{i+1}}{(P_i - P_{i+1})^2} - \frac{\lambda_i}{P_i^2} - \frac{\mu_i}{(1 - P_i)^2} \quad (i = 1, 2, \cdots, m - 1)$$
$$D_{mm} = -\frac{\delta_m}{(P_{m-1} - P_m)^2} - \frac{\lambda_m}{P_m^2} - \frac{\mu_m}{(1 - P_m)^2} \tag{3.2}$$
$$D_{i+1,i} = D_{i,i+1} = \frac{\delta_{i+1}}{(P_i - P_{i+1})^2} \quad (i = 1, 2, \cdots, m - 1)$$
$$D_{ij} = 0, \quad \text{for} \quad |i - j| \geq 2.$$

In Lemma A2 of the appendix, we show that $\mathbf{D}$ is negative definite. Hence all solutions of (3.1) yield maxima. But $L$ is a continuous function for $1 > P_1 > \ldots > P_m > 0$ and so if $L$ has two maxima, there must be a minimum between them. There are no minima and so $L$ has a unique maximum. This completes the proof.

As in [7] and [15], the maximum likelihood estimators are consistent when $m$ and the time points $\{t_i; 1 \leqslant i \leqslant m\}$ remain fixed while the number of items observable at age $t_i$ tends to infinity for all $i$. In this case, the problem is essentially one of estimating the parameters of a multinomial distribution and consistency follows by the law of large numbers. A more interesting problem is the one of consistency when $m \to \infty$ and the widths of the intervals become arbitrarily small—it is hoped to treat this case in a later paper.

The Fisher information matrix $\mathbf{J}$ is the matrix $-\mathbf{D}$, where $\mathbf{D}$ is given by (3.2). $\mathbf{J}$ is a symmetric Jacobi matrix, i.e., it is of the form:

$$\begin{bmatrix} c_1 & d_1 & 0 & 0 & \cdots & 0 & 0 \\ d_1 & c_2 & d_2 & 0 & \cdots & 0 & 0 \\ 0 & d_2 & c_3 & d_3 & \cdots & 0 & 0 \\ 0 & 0 & d_3 & c_4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & c_{m-1} & d_{m-1} \\ 0 & 0 & 0 & 0 & \cdots & d_{m-1} & c_m \end{bmatrix}$$

where $c_i = -D_{ii}$ $(1 \leq i \leq m)$ and $d_i = -D_{i,i+1}$ $(1 \leq i \leq m - 1)$. The inverse $\mathbf{V}$ of $\mathbf{J}$ is a Green's matrix (see, e.g., [16, Ch. 3.3]) and is given by:

$$V_{ij} = a_{\min(i,j)} \cdot b_{\max(i,j)},$$

where

$$a_i = \frac{(-1)^i}{\det (J)} J \begin{pmatrix} 1, 2, \cdots, i - 1 \\ 1, 2, \cdots, i - 1 \end{pmatrix} d_i \cdots d_{m-1} \quad (2 \leq i \leq m - 1)$$

$$b_j = (-1)^j J \begin{pmatrix} j + 1, \cdots, m \\ j + 1, \cdots, m \end{pmatrix} \frac{1}{d_j d_{j+1} \cdots d_{m-1}} \quad (1 \leq j \leq m - 1)$$

$$a_1 = -d_1 d_2 \cdots d_{m-1} / \det(J)$$

$$a_m = (-1)^m J \begin{pmatrix} 1, 2, \cdots m - 1 \\ 1, 2, \cdots m - 1 \end{pmatrix} \Big/ \det (J)$$

$$b_m = (-1)^m,$$

and $J\binom{i_1 i_2 \cdots i_r}{j_1 j_2 \cdots j_r}$ represents the determinant of the matrix formed from $\mathbf{J}$ by removing all rows and columns except rows $i_1, i_2, \cdots, i_r$ and columns $j_1, j_2, \cdots, j_r$.

Denote $\mathbf{V}(\hat{\mathbf{P}})$ as the value of the matrix $\mathbf{V}$ when the $\{P_i\}$ are replaced by their maximum likelihood estimates $\{\hat{P}_i\}$. Then $\mathbf{V}(\hat{\mathbf{P}})$ is an asymptotically unbiased estimate of the variance-covariance matrix of the $\{\hat{P}_i\}$. Thus confidence sets and tests of hypotheses concerning the $\{P_i\}$ can now be constructed. In fact, for testing a null hypothesis, variances should be based on the null situation (see [20, p. 490]); however, these values will be close for large sample sizes. For the numerical example given in the tabulation in Section 2, the matrix $\mathbf{V}(\hat{\mathbf{P}})$ is (only the upper triangular part is shown):

$$\begin{bmatrix} 7.50 & 3.42 & 2.28 & 0.91 \\ & 5.98 & 3.98 & 1.60 \\ & & 5.05 & 2.02 \\ & & & 2.58 \end{bmatrix} \times 10^{-3}.$$

## 4. DISCUSSION AND CONCLUSIONS

The idea of self-consistency was first proposed by Efron [8, Sec. 7], who showed that a similar invariance property held for the PL estimates in the case of single censoring. This article extends that idea to derive an iterative procedure for obtaining maximum likelihood estimates when there is double censoring. This procedure is simpler than the usual one of solving the maximum likelihood iterative equations, which involves calculation and updating of the matrix of second derivatives of $L$ at each stage. (For a discussion, see [20].)

The method of Section 2 is designed for the situation where there is a discrete time scale or the data can be grouped naturally. In this case we need only characterize $F$ by its cumulative probability at the finite number of points which are of interest. If there is a continuous time scale there is clearly some information lost by grouping the data, although this loss is small in practice. Other methods which might be appropriate here are based on estimation of the hazard rate, which is usually assumed constant within each interval (see, e.g., [13, 2, Sec. 3]). The assumptions concerning the timing of the losses and late entries correspond with those made by Kaplan and Meier [15, Sec. 1.4] and will be valid if the inspection procedure is as follows: Examine a cohort all of age $t$, observe the number of deaths $\delta$ since the last examination, record the number of late entries $\mu$ known only to have died at or before $t$; finally lose contact with a number $\lambda$ of the survivors. This applied in the Leiderman study, for example. However, Kaplan and Meier (Sect. 4.1) show how to treat alternative situations, and their modifications of the PL method can be carried over to the procedure of Section 2.

Finally, it is an interesting exercise to generalize the procedure to derive the self-consistent estimates for the $\{P_i\}$ in the problem considered by Harris, Meier and Tukey [13] and also by Mantel [19]. Here, the age of death is never known exactly, but is known only to fall into some interval, perhaps semi-infinite, where this interval differs from item to item.

## APPENDIX

Here we prove two lemmas needed in the proof of the theorem of Section 3. The notation used is the same as that in Sections 2 and 3.

*Lemma A1*: Let $\mathrm{P} = (P_1, P_2, \cdots, P_m)$ be the self-consistent estimates defined by (2.1). (We shall omit the caret signs.) Then P satisfies the likelihood equations (3.1).

*Proof*: First note that if $\lambda_m, \delta_1, \cdots, \delta_m$ are positive then $1 > P_1 > \cdots > P_m > 0$. It is required to prove $\partial L/\partial P_i = 0$ for $i = 1, 2, \cdots, m$. We do this first for $i = m$ and then proceed by induction to show it is true for $i = m - 1, m - 2, \cdots, 1$.

Now

$$P_m = P_{m-1} \cdot \frac{\lambda_m}{\lambda_m + \delta_m'},$$

where

$$\delta_m' = \delta_m + \frac{P_{m-1} - P_m}{1 - P_m} \cdot \mu_m.$$

Therefore, substituting for $\delta_m'$, we obtain:

$$\lambda_m P_{m-1} = (\delta_m + \lambda_m) P_m + (P_{m-1} - P_m)\mu_m P_m/(1 - P_m)$$

or

$$\frac{\delta_m}{P_{m-1} - P_m} - \frac{\lambda_m}{P_m} + \frac{\mu_m}{1 - P_m} = 0.$$

Hence $\partial L/\partial P_m = 0$. For fixed $i < m$ assume $\partial L/\partial P_j = 0$ for $j = m, m - 1, \cdots, i + 1$. Now

$$P_i = P_{i-1} \cdot \frac{n_i' - \delta_i'}{n_i'} = P_{i-1} \cdot \frac{n_{i+1}' + \lambda_i}{n_{i+1}' + \lambda_i + \delta_i'}.$$

Substituting for $\delta_i' = \delta_i + (P_{i-1} - P_i) \sum_{j=i}^{m} (\mu_j/1 - P_j)$ and rearranging terms we obtain:

$$\frac{n_{i+1}' + \lambda_i}{P_i} - \frac{\delta_i}{P_{i-1} - P_i} - \sum_{j=i}^{m} \frac{\mu_j}{1 - P_j} = 0. \qquad \text{(A.1)}$$

By the induction hypothesis, $\sum_{j=i+1}^{m} (\partial L/\partial P_j) = 0$, or

$$-\frac{\delta_{i+1}}{P_i - P_{i+1}} + \sum_{j=i+1}^{m} \frac{\lambda_j}{P_j} - \sum_{j=i+1}^{m} \frac{\mu_j}{1 - P_j} = 0.$$

Substituting in (A.1), we obtain:

$$-\frac{\delta_i}{P_{i-1} - P_i} + \frac{\delta_{i+1}}{P_i - P_{i+1}} + \frac{\lambda_i}{P_i} - \frac{\mu_i}{1 - P_i} + \frac{n_{i+1}'}{P_i} - \sum_{j=i+1}^{m} \frac{\lambda_j}{P_j} = 0. \qquad \text{(A.2)}$$

We now claim that, for $0 \le i \le m - 1$:

$$\frac{n_{i+1}'}{P_i} - \sum_{j=i+1}^{m} \frac{\lambda_j}{P_j} = 0. \qquad \text{(A.3)}$$

Since $P_m = P_{m-1} \cdot \lambda_m/(\lambda_m + \delta_m')$, we have $n_m'/P_{m-1} = \lambda_m/P_m$ and (A.3) is true for $i = m - 1$. Assume (A.3) is true for $i = t$. Then $\sum_{j=t}^{m} (\lambda_j/P_j) = (n_{t+1}' + \lambda_t)/P_t = (n_t' - \delta_t')/P_t = n_t'/P_{t-1}$. Thus (A.3) is true for $i = t - 1$ and by induction true for all $0 \le i \le m - 1$.

Combining (A.2) and (A.3) we obtain:

$$\frac{\partial L}{\partial P_i} = -\frac{\delta_i}{P_{i-1} - P_i} + \frac{\delta_{i+1}}{P_i - P_{i+1}} + \frac{\lambda_i}{P_i} - \frac{\mu_i}{1 - P_i} = 0.$$

The proof of Lemma A1 now follows by induction.

*Lemma A2*: The matrix $\mathbf{D}$ given by (3.2) is negative definite.

*Proof*: We will show that the matrix $\mathbf{J} = -\mathbf{D}$ is positive definite. For $j = 1, 2, \cdots m$, let $J_j$ denote the value of the $j$'th leading principal minor. It suffices to show that $J_1, J_2, \cdots, J_m$ are all positive.

For $1 \le i \le m$, let $x_i = \delta_i/(P_{i-1} - P_i)$, and

$$y_i = (\lambda_i/P_i^2) + [\mu_i/(1 - P_i)^2].$$

Note that $y_i \ge 0$ and $x_i > 0$ since all $\delta_i > 0$ under the hypotheses of the theorem. Then by (3.2),

$$J = \begin{bmatrix} z_1 & -x_2 & 0 & \cdots & 0 \\ -x_1 & z_2 & -x_3 & \cdots & 0 \\ 0 & -x_3 & z_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & z_m \end{bmatrix}$$

where $z_i = x_{i+1} + x_i + y_i$ for $1 \le i \le m$ and $x_{m+1} = 0$.

Thus,

$$J_1 = x_2 + x_1 + y_1 > 0$$

$$J_2 = (x_3 + x_2 + y_2)(x_2 + x_1 + y_1) - x_2^2 > 0$$

and

$$J_i = (x_{i+1} + x_i + y_i)J_{i-1} - x_i^2 J_{i-2}. \qquad \text{(A.4)}$$

If we set $J_0 = 1$ and $J_{-1} = 0$ then (A.4) holds for $i = 1, 2, \cdots m$. We proceed by induction. Assume $J_1, J_2, \cdots, J_{i-1}$ are all positive. Then using (A.4) we have:

$$J_i > x_i J_{i-1} - x_i^2 J_{i-2} = x_i(x_{i-1} J_{i-2} - x_{i-1}^2 J_{i-3} - x_i J_{i-2})$$
$$> x_i x_{i-1}(J_{i-2} - x_{i-1} J_{i-3}),$$

since $J_{i-2} > 0$, $x_{i-1} > 0$.

Iterating, we have:

$$J_i > x_i x_{i-1} \cdots x_1 (J_0 - x_0 J_{-1}) = x_i x_{i-1} \cdots x_1 > 0.$$

The proof of Lemma A2 now follows by induction.

[*Received July 1972. Revised April 1973.*]

## REFERENCES


[1] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E., "An Empirical Distribution Function for Sampling with Incomplete Information," *Annals of Mathematical Statistics*, 26 (December 1955), 641–7.

[2] Barlow, R.E., "Some Recent Developments in Reliability Theory," *Selected Statistical Papers*, Mathematical Centre, Amsterdam, 2 (1968), 49–60.

[3] ———, Bartholomew, D.J., Bremner, J.M. and Brunk, H.D., *Statistical Inference under Order Restrictions*, New York: John Wiley and Sons, Inc., 1972.

[4] Bayley, N., *Bayley Scales for Infant Development*, New York: Psychological Corp., 1969.

[5] Berkson, J. and Gage, R.P., "Calculation of Survival Rates for Cancer," *Proceedings of the Staff Meetings of the Mayo Clinic*, 25 (1950), 270–86.

[6] Boag, J.W., "Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy," *Journal of the Royal Statistical Society*, Ser. B., 11, No. 1 (1949), 15–53.

[7] Brunk, H.D., "Maximum Likelihood Estimates of Monotone Parameters," *Annals of Mathematical Statistics*, 26 (December 1955), 607–16.

[8] Efron, B., "The Two-Sample Problem with Censored Data," *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 4, 1967, 831–53.

[9] Gehan, E.A., "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly Censored Samples," *Biometrika*, 52, Nos. 1 and 2 (1965), 203–23.

[10] ———, "A Generalized Two-Sample Wilcoxon Test for Doubly Censored Data," *Biometrika*, 52, Nos. 3 and 4 (1965), 650–3.

[11] Gilbert, J.P., "Random Censorship," Unpublished Ph.D. thesis, University of Chicago, 1962.

[12] Halperin, M., "Extension of the Wilcoxon-Mann-Whitney Test to Samples Censored at the Same Fixed Point," *Journal of the American Statistical Association*, 55 (March 1960), 125–38.

[13] Harris, T.E., Meier, P. and Tukey, J.W., "Timing of the Distribution of Events Between Observations," *Human Biology*, 22 (December 1950), 249–70.

[14] Herman, R.J. and Patell, R.K.N., "Maximum Likelihood Estimation for Multi-Risk Model," *Technometrics*, 13 (May 1971), 385–96.

[15] Kaplan, E.L. and Meier, P., "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53 (June 1958), 457–81.

[16] Karlin, S., *Total Positivity I*, Stanford University Press, 1968.

[17] Leiderman, P.H., Babu, D., Kagia, J., Kraemer, H.C. and Leiderman, G.F., "African Infant Precocity and Some Social Influences During the First Year," *Nature*, 242 (March 1973), 247–9.

[18] Mantel, N., "Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration," *Cancer Chemotherapy Reports*, 50 (March 1966), 163–70.

[19] ———, "Ranking Procedures for Arbitrarily Restricted Observations," *Biometrics*, 23 (March 1967), 65–78.

[20] ——— and Myers, M., "Problems of Convergence of Maximum Likelihood Iterative Procedures in Multiparameter Situations," *Journal of the American Statistical Association*, 66 (September 1971), 484–91.

[21] Moeschberger, M.L. and David, H.A., "Life Tests under Competing Causes of Failure and the Theory of Competing Risks," *Biometrics*, 27 (December 1971), 909–33.

[22] Peto, R., "Experimental Survival Curves for Interval-Censored Data," *Applied Statistics*, 22, No. 1 (1973), 86–91.